



HAL
open science

Preliminary stage in the development of an artificial intelligence algorithm: variations between 100 surgeons in phase annotation in a video of internal fixation of distal radius fracture

Camille Graëff, Audrey Daiss, Thomas Lampert, Nicolas Padoy, Antoine Martins, Marie-Cécile Sapa, Philippe Liverneaux

► To cite this version:

Camille Graëff, Audrey Daiss, Thomas Lampert, Nicolas Padoy, Antoine Martins, et al.. Preliminary stage in the development of an artificial intelligence algorithm: variations between 100 surgeons in phase annotation in a video of internal fixation of distal radius fracture. *Orthopaedics & Traumatology: Surgery & Research*, 2023, 109 (6), pp.103564. 10.1016/j.otsr.2023.103564 . hal-03976175

HAL Id: hal-03976175

<https://hal.science/hal-03976175v1>

Submitted on 15 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Original article

Preliminary stage in the development of an artificial intelligence algorithm: variations between 100 surgeons in phase annotation in a video of internal fixation of distal radius fracture

Camille Graëff^{1,2}, Audrey Daiss³, Thomas Lampert¹, Nicolas Padoy^{1,2}, Antoine Martins³, Marie-Cécile Sapa³, Philippe Liverneaux^{1,3}

¹ ICube CNRS UMR7357, Strasbourg University, 2-4 Rue Boussingault, 67000 Strasbourg, France

² IHU, Institute of Image-Guided Surgery, Strasbourg, France.

³ Department of Hand Surgery, Strasbourg University Hospitals, FMTS, 1 Avenue Molière, 67200 Strasbourg, France

Corresponding author: Philippe Liverneaux

Department of Hand Surgery, Strasbourg University Hospitals, FMTS, 1 Avenue Molière, 67200 Strasbourg, France

philippe.liverneaux@chru-strasbourg.fr

Abstract

Introduction

In order to be used naturally and widely, an artificial intelligence algorithm of phase detection in surgical videos presupposes an expert consensus defining phases.

Objectives

The aim of the present study was to seek consensus in defining the various phases of a surgical technique in wrist traumatology.

Methods

3,229 surgeons were sent a video showing anterior plate fixation of the distal radius and a questionnaire on the number of phases they distinguished and the visual cues signaling the beginning of each phase. Three experimenters predefined the number of phases (5: installation, approach, fixation, verification, closure) and sub-phases (3a: introduction of plate; 3b: positioning distal screws; 3c: positioning proximal screws) and the cues signaling the beginning of each. The numbers of the responses per item were collected.

Results

Only 216 (6.7%) surgeons opened the questionnaire, and 100 answered all questions (3.1%). Most respondents claimed 5/5 expertise. Number of phases identified ranged between 3 and 10. More than two-thirds of respondents identified the same phase cue as defined by the 3 experimenters in most cases, except for “verification” and “positioning proximal screws”.

Discussion

Surgical procedures comprise a succession of phases, the beginning or end of which can be defined by a precise visual cue on video, either beginning with the appearance of the cue or the disappearance of the cue defining the preceding phase.

Conclusion

These cues need to be defined very precisely before attempting manual annotation of surgical videos in order to develop an artificial intelligence algorithm.

Level of evidence: II

Keywords: phase detection, surgical video, algorithm, artificial intelligence, distal radius fracture

Introduction

One of the very first applications of artificial intelligence (AI) in medicine, in the 1970s, concerned patient-specific antibiotic therapy [1]. Since then, many other applications have been developed, such as algorithms for diagnosis of fractures on standard radiographs [2]. Other algorithms were developed for surgical planning and navigation, especially in neurosurgery, maxillofacial surgery and orthopedic and traumatologic surgery [3]. More recently, studies focused on videos of soft-tissue surgery for phase detection in cholecystectomy [4], sleeve gastrectomy [5], laparoscopic sigmoidectomy [6] and endoscopic myotomy [7]. The precondition for developing this kind of algorithm is to precisely define the phases to be detected. If the algorithms are to be widely used, expert consensus here is indispensable. To our knowledge, there are no phase detection algorithms in orthopedic and traumatologic surgery. In contrast to soft-tissue surgery, techniques here are relatively heterogeneous, making phase definition more difficult. Precise definition of phases and the visual cues marking their start and end enables a protocol to be drawn up for annotation of videos in an annotated database which is essential for machine learning for a specific task.

The present study therefore sought consensus in a panel of surgeons regarding the exact number of phases comprising a wrist trauma procedure and the visual cues marking their boundaries.

The main study hypothesis was that at least two-thirds of the panel would agree on the number of phases seen in a video of internal fixation of wrist fracture. The secondary hypotheses were that at least two-thirds of the panel would agree on the visual cue marking the start of each phase and subphase defined by the experimenters.

Material and methods

A patient provided written consent for the use of a video of his internal wrist fracture fixation.

The material comprised surgeons, a video and a questionnaire. The surgeons were either traumatologists or hand surgeons, with expertise ranging from 1 (non-specialized) to 5 (expert) [8]. The video (video 1) of minimally invasive anterior plate fixation of a distal radius fracture [9] was downloaded from a dedicated platform (qvident®, Caresyntax™, Boston, Massachusetts, United States). It had been recorded using an HD video camera built into the surgical lighting (TruVidia™ Wireless®, TRUMPF Medizin Systeme GmbH + Co. KG™, Saalfeld, Germany) and accelerated at double speed so as to last 8'50''. Three experimenters (2 level 3 and 5 surgeons and 1 AI researcher) decided on a consensual division of the video into 5 phases (figure 1), with 3 subphases in the third. The questionnaire (Table I) concerned the characteristics of the procedure, with 8 single-choice questions on, firstly, the number of phases, and then the visual cues marking the start of each phase (phase beginning either with the appearance of a cue or with the disappearance of the cue defining the previous phase). The questionnaire was progressive, in the sense that it was not possible to go back to a question once the response had been validated. After question 1 (number of phases), the 5 phases determined by the experimenters were made known to the participants. In questions 6, 7 and 8, the 3 subphases determined by the experimenters were made known. The phases

and subphases were made known to ensure that the participants all identified cues at the same time-points in the video.

The method consisted in e-mailing the video and questionnaire to 3,229 surgeons: 2,402 members of a French national orthopedic and traumatologic surgery society, and 827 members of a French national hand surgery society. They were asked to view the video as often as they liked, and to fill out the questionnaire. The questionnaire was sent out on April 1, 2022 and closed on May 20. Analysis consisted in collecting the number of responses to each item of the 8 questions and expressing it as a percentage.

Results

216 of the 3,229 recipients (6.7%) opened the questionnaire; 116 responded partially (3.6%) and 100 completely (3.1%). To avoid bias, only fully completed questionnaires were included for analysis. Table 2 shows the expertise level of the 100 surgeons; most (53%) claimed to be level 5.

The number of phases identified by the respondents ranged from 3 to 10 (figure 2), 10 being the most frequent.

The approach phase began with the “appearance of the knife” for 72% of participants; “appearance of the dermatographic pen” came second (22%). Fixation began with “appearance of the plate + drill guide” for 75% of participants; “appearance of the first K-wire” came second (16%). Verification began with “last appearance of the fluoroscope” for 48% of participants; “disappearance of the screwdriver after positioning the last proximal screw” came second (24%). Closure began with “appearance of the needle holder” for 62% of participants; “appearance of the physiological saline cup” came second (37%). Plate introduction began with “appearance of the plate + drill guide” for 80% of participants; “appearance of the rongeur” came second (14%). Distal screw positioning began with “appearance of the first K-wire” for 68% of participants; “appearance of the first screw” came second (22%). Proximal screw positioning began with “appearance of the first proximal drill guide” for 68% of participants; “appearance of the rongeur at the first proximal orifice” came second (14%); the response predefined by the experimenters came in 4th place, at 6%.

Discussion

Surgery phase detection may be applied in AI contexts or not. In AI, it serves for developing algorithms. Elsewhere, it makes surgeons aware of the fact that the procedure is not experienced in the same way by all. The present study demonstrated this.

The present results need to be discussed quantitatively and qualitatively. Quantitatively, the mean response rates for health professionals in online surveys is $45.8\% \pm 25.0\%$ [10 Meyer]. In the present study, the rate was 6.7%. This may have been because, while the hand surgery society is obviously interested in wrist fracture, the more general orthopedic and traumatologic surgery society is less involved. Another reason could be surgeons' poor knowledge of AI. Slightly less than half of those who opened the questionnaire (46.3%)

responded to all the questions. This may have been due to more than half giving up at the first question because the video was too long. Originally, it lasted 17'40"; although speeded up 2-fold, it was still long, especially if it was to be viewed more than once. Accelerating more than 2-fold was not feasible because of the risk of overlooking short but important sequences. Qualitatively, questionnaire response reliability depended on 2 factors: the participant's expertise and concentration. For 53% of participants, expertise was unquestionable, at level 5; but as much could not be said of the other 47%. Concentration could not be assessed, but might not necessarily match expertise. For all these reasons, the results are to be taken with caution. There were also other study limitations. The response rate was very low (6.7%) and the full-response rate even lower (3.1%). Participants answered questions based on a single video, and it is not sure that our video-1 was representative of this type of surgery, in which there are many variants. And lastly, the definition of the start or end of a phase may differ between individuals: a surgeon has a conceptual point of view, while an AI researcher will select more precise visual cues.

Developing an AI algorithm based on supervised learning includes an indispensable step of manual annotation of data (images, videos, sounds, etc.) with relevant labels (name, type, shape, color, number, etc.) [11]. Expert consensus is essential for homogeneous annotation enhancing the robustness and generalizability of the algorithm [12]. It is therefore essential to define the labels precisely. In the present study of the means of developing an algorithm that could be used by all surgeons, the data in question were the video and the labels were the phases and subphases defined by the experimenters.

Participants varied in their perception of the number of phases. This may have been because no definition of "phase" was provided, and also due to how fine-grained the concept was meant to be, with the idea of "subphases". This goes to show that, before beginning to annotate surgical videos in order to develop an AI algorithm, the phases of the technique need to be defined precisely.

Three type of questionnaire response emerged: identical to the experimenters'; different, but valid in terms of the video; and different and not valid.

About two-thirds of respondents identified the same cues as the experimenters to define the exact start of the approach, fixation, plate introduction, distal screw positioning, and closure.

Some responses differed from the experimenters' but seemed valid in terms of the video. For the approach, a quarter of the respondents chose the appearance of the dermatographic pen. It is not obvious whether tracing the incision is part of the installation or of the approach; moreover, not all surgeons trace the approach before incising, while others trace it before inflating the tourniquet. The appearance of the knife is thus the ideal cue for the start of the approach. For closure, a little over a third chose the appearance of the cup of physiological saline. Saline is used just before the needle holder, but not all surgeons use it at end of procedure and some use it rather during the verification phase. The appearance of the needle holder is thus the ideal cue for the start of closure. For proximal screw positioning, more than two-thirds chose the appearance of the first proximal drill guide. The disappearance of the screwdriver after positioning the last distal screw is followed by the appearance of the first proximal drill guide. However, this sequence does not apply to all procedures, as some surgeons go without any drill guide. The disappearance of the

screwdriver after positioning the last distal screw is thus the ideal cue for the start of proximal screw positioning; the end of one phase may cue the start of the next.

Some deviant responses could not be considered valid. The exact start of the verification phase was variously identified. Oddly enough, although they knew that verification involved fluoroscopy, almost half the respondents chose the last appearance of the fluoroscope. Even so, a quarter chose the disappearance of the screwdriver after proximal screw positioning, in agreement with the experimenters. The start of distal screw positioning was wrongly identified as the appearance of the first screw in slightly less than a quarter of cases; this response cannot be deemed valid, as positioning the first screw is systematically preceded by use of the first K-wire.

Most surgical phase detection algorithms concern non-urgent well-codified procedures such as sleeve gastrectomy [5], laparoscopic sigmoidectomy [6], endoscopic myotomy [7] or cataract [13]. The most widely studied is laparoscopic cholecystectomy, as there are several open-access databases and it is a common procedure with fairly well-defined phases, although some authors report discrepancies in phase number (from 5 to 20) [11]. The present study had the originality of concerning an emergency procedure that is less well codified: distal radius fixation can be performed using many and varied techniques [14-18].

The main study hypothesis, that at least two-thirds of a panel of surgeons would identify the same number of phases in a video of internal fixation of a wrist fracture, was not confirmed. The secondary hypotheses, that at least two-thirds of the panel would identify the start of each phase and subphase identically to the experimenters was confirmed in 5 out of 7 cases, but not for the verification phase or the proximal screw positioning subphase.

From these results, only partly confirming the study hypotheses, it emerges that a surgical procedure comprises successive phases, the start and end of which can be defined by a precise visual cue identifiable on a video. A phase begins either with the appearance of the cue or with the disappearance of the cue defining the previous phase. In conclusion, it is essential to define these cues very precisely before beginning manual annotation in order to develop an AI algorithm in surgery.

Ethical Approval: This study was performed in line with the principles of the Declaration of Helsinki.

Consent to participate: Informed consent was obtained from the patient included in the study.

Consent to Publish: The authors affirm that human research participant provided informed consent for publication of the video 1.

Conflicts of Interest: Philippe Liverneaux has competing interests with Caresyntax. None of the other authors have competing interest.

Funding: This work was partially supported by the Interdisciplinary Thematic Institute HealthTech, as part of the ITI 2021-2028 program of the University of Strasbourg, CNRS and Inserm, by IdEx Unistra (ANR-10-IDEX-0002), SFRI (STRAT'US project, ANR-20-SFRI-0012) and IHU Strasbourg (ANR-10-IAHU-02) under the framework of the French Investments for the Future Program.

Author Contributions: All authors contributed to the study design (Camille Graëff, Audrey Daiss, Thomas Lampert, Nicolas Padoy, Antoine Martins, Marie-Cécile Sapa, Philippe Liverneaux). Material preparation, data collection and analysis were performed by Camille Graëff, Audrey Daiss and Sapa Marie Cecile. The first draft of the manuscript was written by Philippe Liverneaux and Camille Graëff. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

1. Kaul V, Enslin S, Gross SA. History of artificial intelligence in medicine. *Gastrointest Endosc.* 2020 Oct;92(4):807-812. doi: 10.1016/j.gie.2020.06.040. Epub 2020 Jun 18. PMID: 32565184.
2. Kuo RYL, Harrison C, Curran TA, Jones B, Freethy A, Cussons D, Stewart M, Collins GS, Furniss D. Artificial Intelligence in Fracture Detection: A Systematic Review and Meta-Analysis. *Radiology.* 2022 Mar 29:211785. doi: 10.1148/radiol.211785. Epub ahead of print. PMID: 35348381.
3. Vercauteren T, Unberath M, Padoy N, Navab N. CAI4CAI: The Rise of Contextual Artificial Intelligence in Computer Assisted Interventions. *Proc IEEE Inst Electr Electron Eng.* 2020 Jan;108(1):198-214. doi: 10.1109/JPROC.2019.2946993. Epub 2019 Oct 23. PMID: 31920208; PMCID: PMC6952279.
4. T. Czempiel et al., "TeCNO: Surgical Phase Recognition with Multi-Stage Temporal Convolutional Networks," *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, pp. 343–352, 2020.
5. Hashimoto DA, Rosman G, Witkowski ER, Stafford C, Navarette-Welton AJ, Rattner DW, Lillemoe KD, Rus DL, Meireles OR. Computer Vision Analysis of Intraoperative Video: Automated Recognition of Operative Steps in Laparoscopic Sleeve Gastrectomy. *Ann Surg.* 2019 Sep;270(3):414-421. doi: 10.1097/SLA.0000000000003460. PMID: 31274652; PMCID: PMC7216040.
6. Kitaguchi D, Takeshita N, Matsuzaki H, Takano H, Owada Y, Enomoto T, Oda T, Miura H, Yamanashi T, Watanabe M, Sato D, Sugomori Y, Hara S, Ito M. Real-time automatic surgical phase recognition in laparoscopic sigmoidectomy using the convolutional neural network-based deep learning approach. *Surg Endosc.* 2020 Nov;34(11):4924-4931. doi: 10.1007/s00464-019-07281-0. Epub 2019 Dec 3. PMID: 31797047.
7. Ward TM, Hashimoto DA, Ban Y, Rattner DW, Inoue H, Lillemoe KD, Rus DL, Rosman G, Meireles OR. Automated operative phase identification in peroral endoscopic myotomy. *Surg Endosc.* 2021 Jul;35(7):4008-4015. doi: 10.1007/s00464-020-07833-9. Epub 2020 Jul 27. PMID: 32720177; PMCID: PMC7854950.
8. Tang JB, Giddins G. Why and how to report surgeons' levels of expertise. *J Hand Surg Eur Vol.* 2016 May;41(4):365-6. doi: 10.1177/1753193416641590. PMID: 27076602.
9. Liverneaux PA. The minimally invasive approach for distal radius fractures and malunions. *J Hand Surg Eur Vol.* 2018 Feb;43(2):121-130. doi: 10.1177/1753193417745259. Epub 2017 Dec 12. PMID: 29231799.
10. Meyer VM, Benjamens S, Moumni ME, Lange JFM, Pol RA. Global Overview of Response Rates in Patient and Health Care Professional Surveys in Surgery: A Systematic Review. *Ann Surg.* 2022 Jan 1;275(1):e75-e81. doi: 10.1097/SLA.0000000000004078. PMID: 32649458; PMCID: PMC8683255.
11. Garrow CR, Kowalewski KF, Li L, Wagner M, Schmidt MW, Engelhardt S, Hashimoto DA, Kenngott HG, Bodenstedt S, Speidel S, Müller-Stich BP, Nickel F. Machine Learning for Surgical Phase Recognition: A Systematic Review. *Ann Surg.* 2021 Apr 1;273(4):684-693. doi: 10.1097/SLA.0000000000004425. PMID: 33201088.
12. Jin Z. Expert consensus on the colorectal cancer annotation of CT and MRI (2020) (translation). *Chin J Acad Radiol.* 2021;4:141–149. doi: 10.1007/s42058-021-00076-0.
13. S. Morita, H. Tabuchi, H. Masumoto, T. Yamauchi, and N. Kamiura, "Real-Time Extraction of Important Surgical Phases in Cataract Surgery Videos," *Scientific Reports*, vol. 9, no. 1, 2019.
14. Vernet P, Gouzou S, Hidalgo Diaz JJ, Facca S, Liverneaux P. Minimally invasive anterior plate osteosynthesis of the distal radius: A 710 case-series. *Orthop Traumatol Surg Res.* 2020 Dec;106(8):1619-1625. doi: 10.1016/j.otsr.2020.04.024. Epub 2020 Nov 3. PMID: 33153957.
15. Cheng MF, Chiang CC, Lin CC, Chang MC, Wang CS. Loss of radial height in extra-articular distal radial fracture following volar locking plate fixation. *Orthop Traumatol Surg Res.* 2021 Sep;107(5):102842. doi: 10.1016/j.otsr.2021.102842.

16. Toro-Aguilera Á, Martínez-Galarza P, Camacho-Carrasco P, Caballero M, Segur JM. Comparative study between fixed-angle and polyaxial screws in distal radius fixation with two volar locking plates. *Orthop Traumatol Surg Res.* 2021 Sep;107(5):102801. doi: 10.1016/j.otsr.2020.102801.
17. Desclée de Maredsous R, Huet S, Brichoux S, Charissoux JL, Marcheix PS. User satisfaction with single-use instruments versus reusable instruments for treating distal radius fractures with locking volar plates. *Orthop Traumatol Surg Res.* 2022 Jan 31:103217. doi: 10.1016/j.otsr.2022.103217.
18. Ducournau F, Meyer N, Xavier F, Facca S, Liverneaux P. Learning a MIPO technique for distal radius fractures: Mentoring versus simple experience versus deliberate practice. *Orthop Traumatol Surg Res.* 2021 Dec;107(8):102939. doi: 10.1016/j.otsr.2021.102939.

Legends of tables, figures and video

Table 1. Questionnaire sent to 3,229 surgeons: perception of phases in video-1.







Table 2. Level of expertise of the 100 respondents (according to Tang & Giddins).

Figure 1. Phases and subphases according to the 3 experimenters. 1: Installation (incision trace, positioning drill guide on plate, limb cleansing). 2: Approach (from incision to exposure of the distal radius). 3: Fixation (3a: Introduction of plate; 3b: Distal fixation (distal screws); 3c: Proximal fixation (proximal screws)). 4: Verification (reduction under fluoroscopy, range of motion, absence of impingement with flexor pollicis longus). 5: Closure (intradermal suture, dressing). Only these definitions (without the images of phase start) were communicated to the participants.

Figure 2. Responses to the 8 questions by the 100 surgeons completely filling out the questionnaire.

Video 1. Minimally invasive anterior plate fixation of distal radius fracture.

Table 1. Questionnaire sent to 3,229 surgeons: perception of phases in video-1.

1. How many phases do you see in the video?	
1, 2, 3, 4, 5 , 6, 7, 8, 9, 10	
2. When exactly does the approach start?	
Appearance of the operator's hand Appearance of the dermatographic pen Appearance of the Velpeau strip Appearance of the knife Appearance of the rongeur	
3. When exactly does fixation start?	
Appearance of the rongeur Appearance of the plate + drill guide Appearance of the first K-wire Appearance of the first screw Appearance of the fluoroscope for the first time	
4. When exactly does verification start?	
Disappearance of the screwdriver after positioning the last proximal screw Appearance of range of motion verification Appearance of the fluoroscope for the last time Appearance of the surgeon's hand moving the thumb Appearance of the needle holder	
5. When exactly does closure start?	
Appearance of the fluoroscope for the last time Appearance of the cup with the physiological saline Appearance of the needle holder Appearance of the ruler Appearance of the compresses	
6. When exactly does plate introduction start?	
Appearance of the rongeur Appearance of the plate + drill guided Appearance of the first K-wire Appearance of the first screw Appearance of the fluoroscope for the first time	
7. When exactly does distal fixation start?	
Appearance of the first K-wire Appearance of the first screw Disappearance of the screwdriver after positioning the first distal screw Disappearance of the drill guide	

Appearance of the fluoroscope for the first time

8. When exactly does proximal fixation start?

Appearance of the first proximal drill guided

Disappearance of the screwdriver after positioning the second proximal drill guide

Appearance of the rongeur at the first proximal orifice

Disappearance of the screwdriver after positioning the last distal screw

Appearance of the first proximal screw



Responses predetermined by the 3 experimenters (see text) **in bold**. NB: only the questions/responses and not the images were shown to the participants.

Table 2. Level of expertise of the 100 respondents (according to Tang & Giddins).

Level 1-5	Number N	Percentage %
1	15	15
2	13	13
3	8	8
4	11	11
5	53	53

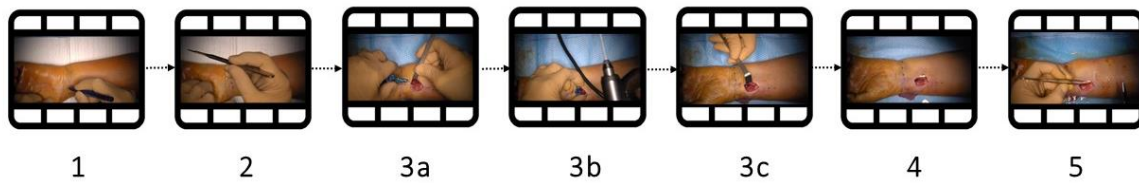
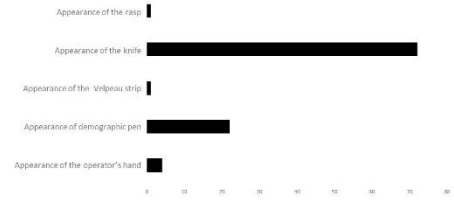


Figure 1. Phases and subphases according to the 3 experimenters. 1: Installation (incision trace, positioning drill guide on plate, limb cleansing). 2: Approach (from incision to exposure of the distal radius). 3: Fixation (3a: Introduction of plate; 3b: Distal fixation (distal screws); 3c: Proximal fixation (proximal screws)). 4: Verification (reduction under fluoroscopy, range of motion, absence of impingement with flexor pollicis longus). 5: Closure (intradermal suture, dressing). Only these definitions (without the images of phase start) were communicated to the participants.

HOW MANY PHASES DO YOU SEE IN THE VIDEO?



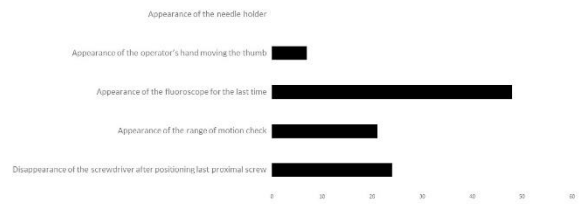
WHEN EXACTLY DOES THE APPROACH START?



WHEN EXACTLY DOES FIXATION START?



WHEN EXACTLY DOES CHECKING START?



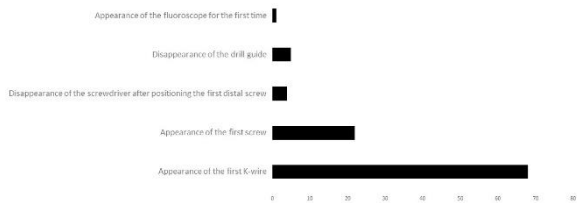
WHEN EXACTLY DOES CLOSURE START?



WHEN EXACTLY DOES PLATE INTRODUCTION START?



WHEN EXACTLY DOES DISTAL FIXATION START?



WHEN EXACTLY DOES PROXIMAL FIXATION START?

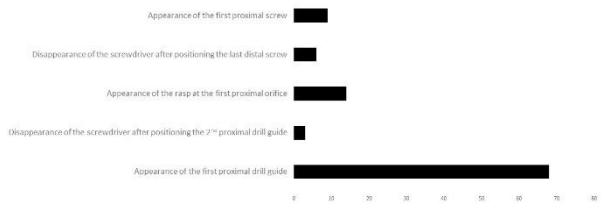


Figure 2. Responses to the 8 questions by the 100 surgeons completely filling out the questionnaire.